

Tipo de artículo: Artículo original

Temática: Inteligencia organizacional y arquitecturas empresariales

Recibido: 02/04/17 | Aceptado: 12/05/17 | Publicado: dd/mm/aa

Mercado de datos de energías renovables

Renewable Energy Data Mart

Manuel Alejandro Fernández Rodríguez¹, Camilo José Tamayo Morales², Olga Yarisbel Rojas Grass³

¹Facultad 3, Universidad de las Ciencias Informáticas, mfrodriguez@estudiantes.uci.cu

²Facultad 3, Universidad de las Ciencias Informáticas, cjtamayo@estudiantes.uci.cu

³Centro de Informatización de Entidades (CEIGE), Facultad 3, Universidad de las Ciencias Informáticas, varisbel@uci.cu

* Autor para correspondencia: mfrodriguez@estudiantes.uci.cu

Resumen

La solución propuesta es un mercado de datos cuyo objetivo fundamental es contribuir al análisis estadístico de los indicadores: capacidad de generación y energía producida de las energías renovables a nivel de país y del nivel mundo. Con el fin de dar cumplimiento al objetivo propuesto, fue centralizada la información que se encuentra almacenada en documentos publicados en sitios como la Oficina Nacional de Estadísticas e Información y la Agencia Internacional de Energías Renovables. Dicho mercado de datos brinda un espacio único donde la información se encuentra centralizada, consistente y disponible. Todo el conjunto de transformaciones que sufren los datos para ser estandarizados y posteriormente visualizados, se llevan a cabo a través de procesos de integración de datos y técnicas de inteligencia de negocios, haciendo uso del conjunto de herramientas de la Suite Pentaho. La metodología que se utiliza es la Propuesta de metodología para el desarrollo de Almacenes de Datos de DATEC, que toma como base el enfoque de Kimball y se ajusta a las condiciones y características de la producción del centro y de la Universidad de las Ciencias Informáticas. Una vez concluidos dichos procesos, se obtiene como resultado un mercado de datos poblado y una aplicación web para la visualización de diferentes reportes que incluyen el análisis de los principales indicadores de las energías renovables.

Palabras clave: centralizada, consistente, disponible, mercado de datos, transformaciones.

Abstract

The proposed solution is a data market whose main objective is to contribute to the statistical analysis of the indicators: generation capacity and energy produced from renewable energies at the country level and the world level. In order to

fulfill the proposed objective, the information stored in documents published at sites such as the National Bureau of Statistics and Information and the International Renewable Energy Agency were centralized. This data market provides a unique space where information is centralized, consistent and available. The whole set of transformations that suffer the data to be standardized and later visualized, are carried out through processes of data integration and business intelligence techniques, making use of the Pentaho suite of tools. The methodology used is the proposed methodology for the development of DATEC Data Warehouses, based on Kimball's approach and adjusted to the conditions and characteristics of the production of the center and the University of Computer Science. Once these processes are concluded, a populated data market and a web application for the visualization of different reports that include the analysis of the main indicators of renewable energies are obtained.

Keywords: *available, centralized, consistent, data mart, transformations.*

Introducción

La información se ha convertido en uno de los recursos más valiosos y la materia prima indispensable para el desarrollo de las empresas, sectores y países; además es determinante en la competitividad y un sustento a los procesos de toma de decisiones. No basta con lograr la persistencia de los datos en el tiempo, es necesario su almacenamiento y procesamiento.

El Centro de Gestión de la Información y Desarrollo de la Energía necesita un sistema de recuperación de información que permita la localización e integración de datos e información procedente de organismos e instituciones que se dedican a producir datos e informes relacionados con las fuentes renovables de energía. Este centro necesita conocer el comportamiento en Cuba y el mundo de los indicadores relacionados a la capacidad y producción de energías renovables, así como la cantidad de dispositivos que la generan. Uno de los aspectos importantes a tener en cuenta es la equivalencia que representa la producción de estas energías con respecto al petróleo. Las fuentes de datos utilizadas en este trabajo provienen de sitios como: la Oficina Nacional de Estadísticas e Información y la Agencia Internacional de Energías Renovables que publican de manera anual informes relacionados a estos indicadores.

Con el fin de obtener la mayor información posible, las principales esferas de la sociedad en el mundo han aplicado la tecnología de los Almacenes de Datos (AD). Los mismos constituyen un instrumento eficaz para la dirección de las empresas, pues permiten realizar preguntas, consultas y análisis de los datos según sus necesidades, sin tener que acudir al personal informático de la institución. Particularmente los AD se han convertido en herramientas que ofrecen beneficios para llevar a cabo los diferentes procesos dentro de las organizaciones, constituyendo uno de los principales eslabones para la toma de decisiones.

A partir de la necesidad que se tiene de que la información se encuentre de manera centralizada, consistente y disponible se propone el desarrollo de un mercado de datos cuyo objetivo fundamental es contribuir al análisis estadístico de los indicadores: capacidad de generación y producción de las energías renovables a nivel de país y del nivel mundo.

Materiales y métodos

Para la realización del trabajo se tuvieron en cuenta diferentes aspectos permitiendo su correcto desarrollo y que a continuación se detallan.

Métodos teóricos:

Análisis histórico-lógico: Mediante el estudio del estado del arte de los Almacenes de Datos y los Mercados de Datos, se conoció la evolución histórica y el desarrollo actual de los mismos.

Analítico-Sintético: Se utilizará para distinguir, extraer y unificar los elementos que forman parte del proceso de construcción del mercado de datos e indicadores relacionados con las fuentes de energías renovables.

Entrevista: Se conocerá las funcionalidades que tiene el mercado de datos a desarrollar, a partir de las necesidades del cliente.

Un AD es un sistema computarizado para guardar registros cuya principal finalidad es almacenar información de manera que los usuarios puedan actualizarla y recuperarla. Una vez reunidos los datos de las fuentes se guardan durante mucho tiempo, permitiendo el acceso a datos históricos y proporcionando una mayor facilidad a la hora de realizar consultas en función de la toma de decisiones[1].

Otra definición de AD es: “una colección de información corporativa derivada directamente de los sistemas operacionales y de algunos orígenes de datos externos. Su propósito específico es soportar la toma de decisiones en un negocio, no las operaciones de un negocio”[2].

Un AD es "una colección de datos, orientados a hechos relevantes del negocio, integrados, que incluyen el tiempo como característica importante de referencia y no volátiles para el proceso de toma de decisiones"[3]. También se puede citar lo planteado por otro reconocido autor como Ralph Kimball (2002), quien define a un Almacén de Datos como: “los AD son una copia de los datos de la transacción estructurados específicamente para la pregunta y el análisis”. Como se ha podido ver, existen disímiles autores y estudiosos del tema que definen de forma muy particular

lo que para ellos representa un almacén de datos. Sin embargo, no deja de ser un hecho que un AD es un sistema para la colección de datos con respecto a temas específicos, en los que la información va a prevalecer de una manera efectiva a lo largo del tiempo, y su principal razón de existir es la de brindar ayuda a una organización o empresa con respecto a la toma de decisiones.

A través del análisis de diversas fuentes existentes sobre la tecnología de AD, muchos autores hacen referencia al término de Mercado de Datos (MD), cuando se refieren al proceso de almacenar información. Josep Curto un estudioso en el ámbito de los AD y MD, define a un MD como: “un subconjunto de datos del Almacén de Dato, con el objetivo de responder a un determinado análisis, función o necesidad y con una población de usuarios específica”[4]. Otros autores definen un MD como: “bases de datos multidimensionales orientadas a una materia específica”[2]. Los MD son un subconjunto de datos de un AD donde se almacenan la mayoría de las actividades de análisis que en el entorno de Inteligencia de negocios se llevará a cabo [5].

A partir del estudio realizado se decide implementar un MD permita analizar los indicadores de capacidad y producción de energías renovables, para ello se definió utilizar la Propuesta de metodología para el desarrollo de Almacén de Datos en DATEC. Esta metodología toma como base el enfoque de Kimball y se ajusta a las condiciones y características de la producción de la Universidad.

La metodología de Kimball incluye los siguientes elementos a tener en cuenta en la creación de un MD:

- Crea los conceptos de Hechos y Dimensiones, lo que indudablemente es muy eficaz en el proceso de la toma de decisiones y proporciona mayor agilidad en el proceso de desarrollo.
- Propone ir construyendo el AD a través de la construcción de los MD departamentales, lo que constituye una estrategia buena y coincide con la división lógica de las empresas, entidades, organismos, entre otros.

Es una metodología madura y reconocida por el resto de la comunidad dedicada al tema. Tiene bien definidas las etapas, actividades, artefactos y roles[6].

Durante el ciclo de vida de esta metodología se destacan las siguientes fases de desarrollo:

- Estudio preliminar o planeación: se realiza el estudio de la entidad cliente, la planeación del proyecto, se definen los objetivos, el alcance preliminar, los costos estimados y otras actividades.

- Requisitos: se realiza en dos direcciones, una, mediante la identificación de las necesidades de información y reglas del negocio; y la otra con un levantamiento detallado de las fuentes de datos a integrar. Después se procede a la definición de los requerimientos.
- Arquitectura y diseño: se definen las estructuras de almacenamiento, se diseñan las reglas de extracción, transformación y carga, definiéndose la arquitectura de información que regirá el desarrollo de la solución.
- Implementación: se diseña físicamente el repositorio de datos, se crean las estructuras de almacenamiento, el área temporal de almacenamiento, se ejecutan las reglas de ETL y se configuran e implementan las herramientas de inteligencia de negocios para la obtención de los elementos que se acordaron con el cliente final.
- Prueba: se realizan las pruebas al sistema desde las Pruebas de Unidad hasta las de aceptación con el cliente final.
- Despliegue: se realiza un despliegue piloto en el cual se configuran los servidores y se instalan las herramientas y se carga una muestra de los datos para demostrar que el sistema funciona. Posterior a la aceptación del cliente se realiza la carga de los datos, así como la capacitación y transferencia tecnológica.
- Soporte y mantenimiento: tras la implantación de la solución se brindan los servicios de soporte en línea, vía telefónica, web u otras según el contrato firmado y las condiciones de soporte establecidas.
- Gestión y administración del proyecto: a lo largo del ciclo de vida se realizan actividades de control, gestión y chequeo del desarrollo, los gastos, las utilidades, los recursos y demás actividades por parte del grupo de dirección del proyecto[6].

El MD desarrollado no transita por todas las fases del ciclo de vida de la metodología, solo llega hasta la fase de pruebas. Atendiendo a que las dos fases posteriores a esta son ejecutadas por otras personas.

El proceso de ETL se utiliza para la construcción de los MD, ya que es necesario que la información contenida represente la realidad del negocio, sea confiable y esté disponible en el momento que los usuarios y la organización la necesiten. Para los procesos ETL, es preciso contar con herramientas que permitan reducir tiempo y costos. El desarrollo y la diversificación de las herramientas para los procesos de ETL, actualmente son crecientes, y se refleja en la amplia variedad de herramientas tanto comerciales como de código abierto[7]. Las herramientas que se utilizan

para la integración de los datos es Pentaho Data Integration para realizar la ETL y el Datacleaner para la limpieza de los datos.

- Pentaho Data Integration 4.2.0 es una herramienta multiplataforma que permite extraer la información de las diferentes fuentes, transformar la información a través de un modelo dimensional y cargar los resultados de la transformación en una base de datos tipo AD, a partir del entorno gráfico de desarrollo[7].
- DataCleaner 1.5.4 permite la evaluación del nivel de calidad de los datos contenidos en el sistema de información. Es una aplicación muy fácil de usar, genera sofisticados informes y gráficos que permiten a los usuarios determinar de un vistazo el nivel de calidad de los datos, identificar y analizar la estructura del origen de datos y combinar resultados y gráficos, creando vistas fáciles de interpretar para evaluar la calidad de los datos[8].

Las herramientas de inteligencia de negocios han sido creadas para ayudar a la toma de decisiones entre las diferentes empresas e instituciones de un estado. Además, permiten mostrar una visión general de todos los procesos de la entidad a sus directivos, facilitando un mejor entendimiento en el análisis y en la presentación de los datos.

Las tecnologías de BI son herramientas de soporte de decisiones que permiten en tiempo real, acceso interactivo, análisis y manipulación de información crítica para la empresa. Intentan ayudar a las personas a entender los datos más rápido, a fin de que puedan tomar mejores y más rápidas decisiones y finalmente, mejorar sus movimientos hacia la consecución de objetivos de negocios. A continuación, se mencionan las herramientas definidas para realizar la inteligencia de negocio.

Pentaho Schema Workbench 3.2.1 es una interfaz de diseño que permite crear y probar esquemas de cubos Mondrian OLAP visualmente. La Plataforma de BI de Pentaho incrusta el motor de consulta Mondrian, como parte de su arquitectura. Además, permite la ejecución de consultas. Está distribuida bajo la licencia EPL[9].

Pentaho BI Server 4.5.0 es una herramienta de reporte que permite crear informes, ya sea para ejecutarlos directamente o para publicarlos en la plataforma de Inteligencia de negocios y que desde allí pueden ser utilizados por los usuarios. Incluye asistentes para facilitar la configuración de propiedades, cuenta también con un editor de consultas para facilitar la confección de los datos que serán utilizados en un informe. [10]

Resultados y discusión

La construcción del MD de energías renovables constituye una solución viable a la problemática planteada, teniendo en cuenta que permite el almacenamiento de la información y el análisis de los indicadores de cantidad de dispositivos, capacidad de energía y cantidad de energía producida. Para guiar el proceso de desarrollo de la solución se definió la Metodología para el Desarrollo de AD de DATEC. El diseño se realizará según el esquema constelación de hechos, como modo de almacenamiento se utilizará ROLAP y la suite de Pentaho para el proceso de inteligencia de negocio y extracción, transformación y carga de los datos.

Definición de la arquitectura de un mercado de datos

La arquitectura base de un MD define los aspectos arquitectónicos de la solución. Esta presenta cuatro niveles:

- Fuente de datos: se refiere al origen de los datos.
- Subsistema de integración: incluye los procesos que permiten que los datos de las fuentes sean extraídos, transformados y cargados hacia las fuentes destino.
- Subsistema de almacenamiento: base de datos relacional que contiene las tablas de dimensiones y hechos cargadas a través de los procesos de ETL.
- Subsistema de visualización: comprende las interfaces orientadas a usuario que extraen información y le facilitan la toma de decisiones.

Implementación de los subsistemas

En esta etapa se realiza la implementación de cada uno de los subsistemas que conforman el MD: subsistemas de almacenamiento, integración y visualización. Una vez concluido el proceso de construcción del Mercado de datos de energías renovables, se hace necesaria la validación de la solución mediante la aplicación de técnicas y pruebas.

Para la implementación del subsistema de almacenamiento fue necesaria la utilización de un gestor de base de datos. El SGBD utilizado fue PostgreSQL que permitió representar el diseño de la solución, mediante la creación de las tablas con sus respectivos atributos y relaciones. La definición del modelo de datos se representa en la siguiente figura:

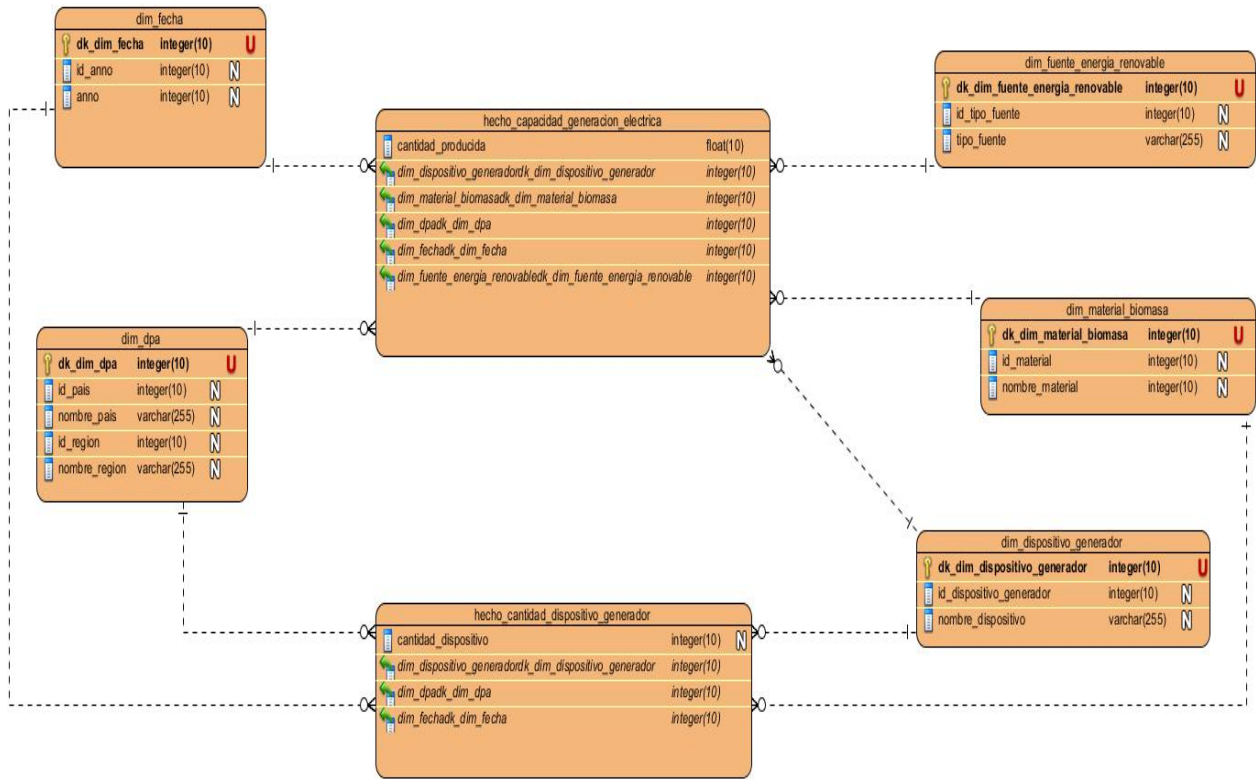


Figura1: Modelo entidad-relación del MD representado los hechos y las dimensiones.

Los esquemas y las tablas permiten brindar una idea general de la base de datos, así como analizar y profundizar los contenidos básicos de la misma. Para la solución propuesta se cuenta con siete tablas identificadas. Para una buena organización de las mismas se agruparon por esquemas, las dimensiones están contenidas dentro del esquema “dimensiones” y las tablas del hecho está contenidas en el esquema “hecho”.

Para la implementación del subsistema de integración se realiza el proceso de integración de los datos que consta de tres etapas fundamentales relacionadas entre sí: extracción, transformación y carga de los datos. Para identificar y corregir los problemas se realiza la limpieza de los datos que permite llenar valores ausentes y corregir errores. Una vez que los datos son transformados se cargan, poblando las dimensiones y los hechos que conforma la estructura del subsistema de almacenamiento del Mercado de datos de energías renovables.

VALORACIÓN ECONÓMICA Y APOORTE SOCIAL

El mercado de datos de manera general permite realizar análisis sobre los datos estadísticos existentes sobre las energías renovables. Permite conocer como se ha ido comportando el crecimiento de utilización de dispositivos que se

utilizan con el fin de generar más energías, así como su producción. También se incluyen los beneficios que trae con respecto al tema del medio ambiente ya que son energías limpias que no contaminan el aire, además se conoce la equivalencia en petróleo de las energías renovables producidas. De manera social garantiza la integridad, consistencia y disponibilidad de toda la información trayendo consigo la reducción de los tiempos empleados para realizar análisis que anteriormente se realizaban manualmente.

Conclusiones

El trabajo realizado cumple con el objetivo planteado inicialmente, que es desarrollar un Mercado de datos de energía renovable que permita la integración, consistencia y disponibilidad de los datos publicados en diferentes fuentes. Para el desarrollo de este mercado de datos se utilizó la metodología de DATEC, el modo de almacenamiento ROLAP y el modo de representación en esquema constelación de hechos. Se obtuvo el modelo de datos, el diseño de los cubos OLAP y las transformaciones a partir de los requisitos de información, funcionales y no funcionales definidos. La solución implementó los subsistemas de integración, almacenamiento y visualización permitiendo el análisis de los indicadores: capacidad de generación y energía producida de las energías renovables a nivel de país y a nivel global.

Referencias

- [1] **Aguilar, L.J.** *Estado actual de las tecnologías data warehousing y OLAP aplicadas a bases de datos espaciales* IV SIMPOSIO INTERNACIONAL DE SISTEMAS DE INFORMACIÓN E INGENIERIA DE SOFTWARE EN LA SOCIEDAD DEL CONOCIMIENTO, 2006. II.
- [2] **Michael Corey, M.A., Ian Abramson, Ben Taub** *Oracle Data Warehousing*. 2002.
- [3] **Inmon, W.H.**, *Using the Data Warehouse*. 1997.
- [4] **Standen, J.** *Data warehouse or Data Mart*. 2008.
- [5] **Claudia Imhoff, N.G.**, Jonathan Geiger, *Mastering Data Warehouse Design, Relational and Dimensional Techniques*. 2003.
- [6] **Hernández, I.Y.**, *PROPUESTA DE METODOLOGIA PARA EL DESARROLLO DE AD EN DATEC*. 2010: La Habana.

- [7] **Corporation., P.P. Powerful Analytics Made Easy.** 2014; Available from: <http://www.pentaho.com/explore/pentaho-data-integration/>.
- [8] **DataCleaner.** The premier data quality solution. 2014; Available from: <http://datacleaner.org/>.
- [9] **mysqlworkbench.** Schema_workbench. 2014; Available from: <http://mysqlworkbench.org/category/tutorial/>.
- [10] **Marcos Patricio Armas Villamarin, S.D.L.C. DESARROLLO E IMPLEMENTACIÓN DE APLICACIONES DE SUPERVISIÓN Y CONTROL DE GENERACIÓN ELÉCTRICA Y ANÁLISIS DE DATOS HISTÓRICOS PARA LA EMPRESA ELÉCTRICA PROVINCIAL COTOPAXI S.A.** 2012; Available from: <http://repositorio.espe.edu.ec/bitstream/21000/5965/1/T-ESPE-034443.pdf>.